FVE-novel: Recovering Draft Genomes of Novel Viruses and Phages in Metagenomic Data

Saima Sultana Tithi1, Frank O. Aylward2, Roderick V. Jensen2, Liqing Zhang1\*

1Department of Computer Science, Virginia Tech, Blacksburg, VA, 24061

2Department of Biological Sciences, Virginia Tech, Blacksburg, VA, 24061

\*Address for Correspondence: [lqzhang@vt.edu](mailto:lqzhang@vt.edu)

Supplementary materials:

|  |  |  |  |
| --- | --- | --- | --- |
| **Station ID** | **SRA ID** | **Paired-end reads** | **Time points (hr)** |
| Station 6 | SRX2912968 | 18,893,711 | 0 |
| Station 14 | SRX2912972 | 13,642,145 | 16 |
| Station 18 | SRX2912964 | 21,075,160 | 32 |
| Station 22 | SRX2912992 | 29,065,983 | 48 |
| Station 28 | SRX2912996 | 19,841,435 | 64 |
| Station 32 | SRX2912975 | 25,336,967 | 80 |
| Station 37 | SRX2912979 | 20,041,567 | 96 |
| Station 52 | SRX2912983 | 12,325,784 | 112 |
| Station 56 | SRX2912998 | 29,996,390 | 128 |
| Station 61 | SRX2913002 | 26,934,921 | 144 |
| Station 67 | SRX2912985 | 15,295,998 | 160 |
| Station 70 | SRX2912986 | 18,471,506 | 172 |

Suppl. Table : Description of the 12 viral-metagenomic samples collected from the study Aylward at el. Along with the time points when these samples were collected, as the sample at station 6 was collected first, we considered the time for collecting this sample as 0 hour.



Suppl. Figure 1: Comparison of S0 with 153 kbp scaffold representing the dominant strain of the novel virus recovered from S0.



Suppl. Figure 2: Comparison of S1 with 153 kbp scaffold representing the dominant strain of the novel virus recovered from S0.



Suppl. Figure 3: Comparison of S6 with 153 kbp scaffold representing the dominant strain of the novel virus recovered from S0.



Suppl. Figure 4: The log2-scaled depth of coverage of S6 (80 kbp) across 12 ocean samples (showing median coverage of per 1000 bp window with step size 1)

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample** | **Start position (bp)** | **End position (bp)** | **# of strains** |
| 1 | 132,719 | 152,719 | 3 |
| 2 | 27,126 | 47,126 | 3 |
| 3 | 79,335 | 99,335 | 3 |
| 4 | 49,195 | 69,195 | 3 |
| 5 | 114,254 | 134,254 | 2 |
| 6 | 88,301 | 108,301 | 3 |
| 7 | 96,101 | 116,101 | 3 |
| 8 | 71,396 | 91,396 | 2 |
| 9 | 108,715 | 128,715 | 2 |
| 10 | 69,477 | 89,477 | 2 |

Suppl. Table : Result of applying viral haplotype reconstruction tool, TenSQR on 10 randomly selected pieces of 153 kbp dominant strain of the novel virus recovered from S0. Here, each piece is 20 kbp long and for each piece, TenSQR reported multiple strains, it implies that multiple strains of this virus is present in the sample.



Suppl. Figure 5: Comparison of S2 with 151 kbp scaffold representing the extended and complete version of S2.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample** | **Start position (bp)** | **End position (bp)** | **# of strains** |
| 1 | 64,623 | 84,623 | 3 |
| 2 | 76,337 | 96,337 | 4 |
| 3 | 112,158 | 132,158 | 2 |
| 4 | 1,012 | 21,012 | 2 |
| 5 | 5,581 | 25,581 | 2 |
| 6 | 59,837 | 79,837 | 2 |
| 7 | 90,933 | 110,933 | 3 |
| 8 | 131,132 | 151,132 | 2 |
| 9 | 39,143 | 59,143 | 2 |
| 10 | 125,497 | 145,497 | 3 |

Suppl. Table : Result of applying viral haplotype reconstruction tool, TenSQR on 10 randomly selected pieces of 151 kbp scaffold representing the extended and complete version of S2. Here, each piece is 20 kbp long and for each piece, TenSQR reported multiple strains, it implies that multiple strains of this virus is present in the sample.



Suppl. Figure 6: Comparison of S3 with 177 kbp dominant strain of Prochlorococcus phage P-SSM4 (recovered from pieces of S3).



Suppl. Figure 7: Comparison of 177 kbp dominant strain with Prochlorococcus phage P-SSM4.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample** | **Start position (bp)** | **End position (bp)** | **# of strains** |
| 1 | 107,315 | 127,315 | 5 |
| 2 | 23,857 | 43,857 | 3 |
| 3 | 60,539 | 80,539 | 5 |
| 4 | 97,235 | 117,235 | 7 |
| 5 | 10,893 | 30,893 | 4 |
| 6 | 120,655 | 140,655 | 4 |
| 7 | 17,466 | 37,466 | 4 |
| 8 | 150,376 | 170,376 | 5 |
| 9 | 109,998 | 129,998 | 5 |
| 10 | 57,176 | 77,176 | 5 |

Suppl. Table : Result of applying viral haplotype reconstruction tool, TenSQR on 10 randomly selected pieces of 177 kbp dominant strain of Prochlorococcus phage P-SSM4 (recovered from pieces of S3). Here, each piece is 20 kbp long and for each piece, TenSQR reported multiple strains, it implies that multiple strains of this virus is present in the sample.



Suppl. Figure 8: Comparison of S8 with 183 kbp dominant strain of Prochlorococcus phage P-HM2 (recovered from pieces of S8).



Suppl. Figure 9: Comparison of 183 kbp dominant strain with Prochlorococcus phage P-HM2.

|  |  |  |  |
| --- | --- | --- | --- |
| **Sample** | **Start position (bp)** | **End position (bp)** | **# of strains** |
| 1 | 161,965 | 181,965 | 3 |
| 2 | 8,470 | 28,470 | 3 |
| 3 | 37,142 | 57,142 | 3 |
| 4 | 78,461 | 98,461 | 2 |
| 5 | 123,677 | 143,677 | 2 |
| 6 | 30,857 | 50,857 | 4 |
| 7 | 68,632 | 88,632 | 2 |
| 8 | 99,087 | 119,087 | 2 |
| 9 | 117,063 | 137,063 | 3 |
| 10 | 2,356 | 22,356 | 2 |

Suppl. Table : Result of applying viral haplotype reconstruction tool, TenSQR on 10 randomly selected pieces of 183 kbp dominant strain Prochlorococcus phage P-HM2 (recovered from pieces of S8). Here, each piece is 20 kbp long and for each piece, TenSQR reported multiple strains, it implies that multiple strains of this virus is present in the sample.